

Assessment and Testing in Mathematics Education

Christine Suurtamm and Michael Neubrand

Introduction

The purpose of Topic Study Group 33 was to address issues related to assessment in mathematics at all levels and in a variety of forms. Assessment and evaluation play an important role in mathematics education as they often define the mathematics that is valued and worth knowing. Furthermore, sound assessment provides important feedback about students' mathematical thinking that prompts student and teacher actions to improve student learning.

Our Topic Study Group sought contributions of research in and new perspectives on assessment in mathematics education that address issues in current assessment practices. Initially we saw these issues as falling into two main strands, large-scale assessment and classroom assessment. Our original call suggested that papers might address one or more of the following topics:

Organizers Co-chairs: Christine Suurtamm (Canada), Michael Neubrand (Denmark); Team Members: Belinda Huntley (South Africa), Liv Sissel Grønmo (Norway), David C. Webb (USA), Martha Koch (Canada), Heidi Krzywacki (Finland); Liaison IPC Member: Johann Engelbrecht (South Africa).

C. Suurtamm (✉)
University of Ottawa, Ottawa, Canada
e-mail: suurtamm@uottawa.ca

M. Neubrand
Universität Oldenburg, Oldenburg, Germany
e-mail: michael.neubrand@uni-oldenburg.de

Large-Scale Assessment

- Issues related to the development of large-scale assessments, which might include such areas as the conceptual foundations of such assessments, designing tasks that value the complexity of mathematical thinking, etc.
- Issues related to the purposes and use of large-scale assessment in mathematics.
- Issues related to the development of large-scale assessment of mathematics teachers' mathematical and pedagogical content knowledge.

Classroom Assessment

- Issues connected to the development of teachers' professional knowledge of assessment and their use of assessment in the mathematics classroom.
- Issues and examples related to the enactment of classroom practices that reflect current thinking in assessment and mathematics education (e.g. the use of assessment for learning, as learning, and of learning in mathematics classrooms)

Broad Issues

- The development of assessment tasks that reflect the complexity of mathematical thinking, problem solving and other important competencies.
- The design of alternative modes of assessment in mathematics (e.g. online, investigations, various forms of formative assessment, etc.).

We received over 50 papers from a range of countries and continents and needed to solicit assistance from committee members and others to review the papers. All papers were reviewed by at least two reviewers. The papers presented a wide variety of issues in assessment and testing and most of the papers were accepted for plenary presentations, small group presentations or poster presentations. The difficult task for the co-chairs was to create a meaningful schedule so that all of these issues could be presented and discussed within the time frame allotted for the Topic Study Group Sessions.

The work was organized into three strands:

- Strand 1: Large-scale assessment and the implications for the development of teaching and learning
- Strand 2: Classroom assessment and developing students' and teachers' knowledge
- Strand 3: Task and test design: Various perspectives

Papers were then categorized according to these three strands and after our initial meeting to introduce the topics and structure of the group, each day consisted of the presentation of plenary papers or posters that are connected to these three strands, and then a division into three subgroups with one subgroup focused on each strand. We also had a poster session with open discussion within the TSG program, as well as posters shown only in the general poster exhibition. The following presents a summary of the main themes presented and discussed in each of the strands. We have also included the ideas from the plenary papers which typically stretched over several strands.

Strand 1: Large-Scale Assessment and the Implications for the Development of Teaching and Learning

There were over 15 papers and several plenary papers presented in this strand over several days. Numerous issues emerged through the discussion of the papers. The range of papers demonstrates that mathematics education researchers are using large-scale assessment results for many different purposes and to investigate a range of complex aspects of mathematics teaching and learning in various contexts. For instance, there are comparative studies (e.g. Wo, Sha, Wei, Li and An) and studies analyzing issues in special regions (e.g. Cheung; Fengbo; Leung; Mizumarchi), studies concentrating on specific topics (e.g. Hodgen, Brown, Coe, and Küchemann), and studies of a more experimental nature (e.g. Li). Several papers that were presented illustrate the challenge of looking for broader trends or patterns across schools and districts while being careful to acknowledge and investigate local contextual factors.

Other papers discuss the use of assessments to investigate a range of factors such as students' higher order thinking skills at different levels of schooling (e.g. Bai; Zhang), gaps in knowledge (e.g. Gersten and Woodward), teacher knowledge (e.g. Shalem, Sapire and Huntley), or teaching approaches (e.g. Thompson), and to improve connections of instruction, assessment, and learning (e.g. Paek). These papers remind us that great care must be taken to ensure that the interpretations being made from the test scores are appropriate. Paper presentations and discussion suggest that using a range of methodological approaches may help to better address the complex questions being investigated in assessment in mathematics education research. For instance, cluster analysis of large-scale data can be used to find patterns in scores but methods such as case study, think aloud protocols while students respond to test items, student and/or teacher focus groups and interviews would enrich our understanding of the patterns observed. The use of a variety of methods is helpful in making sound assertions from data from large-scale assessments.

Strand 2: Classroom Assessment and Developing Students' and Teachers' Knowledge

The papers in this strand were organized into several different categories for presentation: classroom assessment in primary grades (e.g. Makar, and Fry; Hunsader, Thompson, and Zorin), assessing conceptual thinking in the classroom, and teachers' knowledge (Grønmo, Kaarstein, and Ernest). Specific topics that arose in presentation and through discussion included:

- The teachers' role and conceptions of assessment and mathematics (e.g. Esen, Cakiroglu, and Capa-Aydin; Hoch and Amit)
- Task design to elicit students' thinking (e.g. Kim, Kim, Lee, Joen, and Park)
- The students' role and responses to open, though provoking questions (e.g. Mangulabnam)
- Development of students' self-reflection, self-assessment, and self-regulation (e.g. Teong, and Cheng)
- Developing transparency, for students in particular, in classroom assessment (e.g. Semana and Santos)
- Teachers' experiences in implementing formative assessment (e.g. Koch and Suurtamm; Krzywacki, Koistin, and Lavonen)
- Assessing conceptual understanding through alternative assessments (e.g. Türeğün)

Across all of these categories was a strong emphasis on formative assessment and at the heart of most, if not all of these papers, was the desire by either researcher and/or teachers to make sense of what students are thinking and learning. The presentations attended to various ways that students' mathematical reasoning is elicited and interpreted by teachers through classroom assessment.

There was also a great deal of discussion about initiatives in various jurisdictions to improve classroom assessment and to support teachers' use of formative assessment. These initiatives included assessment resources, collaboration, professional development, and support from Ministries of Education. It was noted that this support coupled with valuing teachers' autonomy and professional judgment seemed to provide fertile ground for sound classroom assessment practices. It was noted however that this is not occurring in all jurisdictions and we discussed the differences in teacher autonomy in different countries. International forums such as ICME provide a rich setting where these comparative discussions can occur and may prompt other jurisdictions to develop new initiatives that support strong classroom assessment.

Strand 3: Task and Test Design, Various Perspectives

The core of test design is the creation of appropriate tasks. However that is a business that requires consciousness about the various purposes tests are constructed for. Therefore, tasks and test design has aspects of conceptual and practical nature, and implementation issues are also to be considered.

Thus, the sessions in this Strand addressed a rich bundle of aspects. We started with reports on studies on teachers' knowledge (e.g. Webb). One question addressed is, how knowledge and behavior come together, and how that interplay can be measured. Studying teachers' knowledge has also internationally comparative aspects, insofar as pedagogical content knowledge for teaching has to be effectively operationalized (e.g. Kaarstein).

How specific items for goals of assessment can be constructed appropriately—closed or open as well—was the topic of the next session, with contributions of Hong and Choi; Toe and de la Torre; Kwong and Ming; Kang and Lee; Hong, Kim, Lee, and Joo. Also this question has various perspectives from elementary mathematics classrooms to college-bound students; various mathematical topics have to be attended from big ideas about measurement to the issues of learning to prove; dealing with the answers of the students is decisive and ranges from descriptions to the analysis of the competencies which can be detected in the student responses by appropriate models. All these aspects require also the discussion of methodological issues.

Finally, we also discussed some broader aspects of using tests. One topic was how teachers view and use an on-line, formative assessment system and what conclusions they can draw for their teaching (e.g. Stacey and Steinle). And even broader, was the general question as to whether entrance tests to universities are necessary (e.g. Kohanova).

Concluding Remarks

There was discussion within this topic study group as to whether it should have been two topic study groups—one for large-scale assessment and another for issues in classroom assessment. The discussion concluded by recognizing that these should not be separated as it is critically important that these two groups share their issues, ideas and practices if there is to an alignment between assessment that is ongoing, such as in a classroom and assessment that is an event, such as in large-scale assessment. The participants also found that discussions across countries pointed up many similarities in issues such as teacher professional development in assessment, transparency to students, task design to elicit student thinking, and meanings given to assessment results.

Acknowledgment The contribution of the committee members, authors and participants in TSG 33 are sincerely acknowledged.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.